Validation Plan

Name of the device and project:

HippoVolume.AI

Intended Use:

Assist the radiologist with labeling and calculating the volume of the hippocampus on brain MRI scans. Tracking the hippocampus volume is an important measure of the progression of brain disorders such as Alzheimer's disease. While a radiologist would have to analyze every slice of the 3D volume (axial, sagittal and coronal) of the MRI scan tracing the shape of the hippocampus, HippoVolume.AI requires much less computation and radiologist intervention to estimate the volume and segment the hippocampus.

Algorithm Description:

The system is an end-to-end AI application which includes a machine learning agent integrated into the clinical viewer. A U-NET architecture is used to automatically segment and predict the volume of the hippocampus from the brain MRI scans that are cropped around the area of interest.

Training dataset information:

The training images come from the publicly available dataset: Medical Decathlon Competition.
It is a collection of NIFTII files, with one file per volume and one file for the corresponding mask (label). The labels were created manually by radiologists, though this labeling may induce errors depending on the years of experience of the radiologist. The original images are T2 MRI Scans which have been cropped using the algorithm HippoCrop.

Validation dataset requirements:

Note: Our algorithm ensures a good accuracy with these characteristics, outside of these ranges we do not guarantee a good performance.

- Patient:
  - age > 20
  - gender: male or female
- Image characteristics:
  - modality: MRI
  - body part examined: Head
- Conditions:
  The patient may have neurodegenerative conditions. brain MRI including tumors are not recommended.

Performance Algorithm

The system outputs a predicted mask for each volume. The system is then evaluated using four metrics (the higher these scores are the better):

1. Dice score: To gauge the similarity between the prediction and true mask.
2. Jaccard score: To measure how much the prediction and true mask intersect.
3. Sensitivity: measures the ability to detect hippocampus pixels
4. Specificity: measures the ability to detect the surroundings from the hippocampus.

Example of results:


```
{
  "volume_stats": [
    {
      "filename": "hippocampus_006.nii.gz",
      "dice": 0.9162573673870335,
      "jaccard": 0.8454566054837979,
      "sensitivity": 0.8752052545155994,
      "specificity": 0.9989216158508092
    },
    {
      "filename": "hippocampus_053.nii.gz",
      "dice": 0.9307463532077609,
      "jaccard": 0.8704635761589404,
      "sensitivity": 0.9337880079568059,
      "specificity": 0.9982706558672728
    },
    {
      "filename": "hippocampus_287.nii.gz",
      "dice": 0.9243792325056434,
      "jaccard": 0.8593913955928646,
      "sensitivity": 0.8854054054054054,
      "specificity": 0.9992424857289722
    },
```


On average, the dice score is > 0.9102 which is a promising result (our algorithm was trained on fewer epochs than expected). The overall jaccard score was 0.82.

Mean sensitivity score: 0.864

Mean specificity score: 0.991

Validation requirements:

On validation, we need to ensure a dice score > 0.9 each time.